A versatile genome-scale PCR-based pipeline for high-definition DNA FISH

Magda Bienko[1–3,7], Nicola Crosetto[1–3,7], Leonid Teytelman[1–3], Sandy Klemm[4], Shalev Itzkovitz[1–3] & Alexander van Oudenaarden[1–3,5,6]

We developed a cost-effective genome-scale PCR-based method for high-definition DNA FISH (HD-FISH). We visualized gene loci with diffraction-limited resolution, chromosomes as spot clusters and single genes together with transcripts by combining HD-FISH with single-molecule RNA FISH. We provide a database of over 4.3 million primer pairs targeting the human and mouse genomes that is readily usable for rapid and flexible generation of probes.

DNA FISH is a widely used method with broad applications, including genetic diagnostics and chromosome architecture analysis[1,2]. Despite continuous advancements, several aspects of this technology require further improvement to enable its potential to be fully exploited in both research and diagnostic laboratories. For example, even though a formidable portfolio of ready-to-use probes is available from various commercial sources, the choice of probes is usually restricted to a relatively small number (<200) of loci, most of which belong to the human genome. Probes targeting other loci or species can be developed from genomic DNA fragments cloned in BACs and fosmids or using array-based technology[3,4]. However, implementing these methods in house or outsourcing them to commercial parties can be challenging, if not impossible, for many laboratories for both logistic and economic reasons. Another important limitation is that flexible selection and control of the pool of DNA fragments constituting a FISH probe, which would enable combinatorial labeling and tailoring of the probe size at the discretion of the user, is technically unfeasible with methods based on cloned DNA fragments and is not cost effective with array technology.

To overcome these limitations and enable a wider community of researchers to fully exploit the potential of DNA FISH, we designed ready-to-use human and mouse genomic libraries of PCR primer pairs with optimal thermodynamic features, delimiting amplicons 200–220 nucleotides (nt) in length. After filtering out primer pairs that amplify multiple targets and crosshybridizing amplicons, we obtained a database consisting of 4,823,784 and 4,387,601 unique primer pairs for the human and mouse genome, respectively (**Fig. 1a**). The database can be accessed at www.hdfish.eu. Over 90% of the human and mouse genomes are densely covered by our primers, with more than 80 primer pairs per 100 kb (**Fig. 1b**,**c** and **Supplementary Fig. 1a**,**b**). Previous attempts to use PCR for unique DNA FISH probe generation were either limited to very few loci or based on time-consuming primer design targeting longer amplicons[5,6]. In contrast, our primer libraries are easy to access and ready to use, allowing highly specific double-stranded probes to be rapidly generated for virtually any desired genomic locus by fluorescently labeling pooled amplicons after PCR with a flexibility and cost effectiveness that is otherwise unachievable with other methods (**Supplementary Fig. 2a**,**b** and **Supplementary Note**).

As a proof of principle, we constructed a probe consisting of 50 200-nt unique fragments obtained by PCR and labeled with the Universal Linkage System[7] targeting the *HER2* (also known as *ERBB2*) locus on chromosome 17. The effective target size (ETS) of this probe (that is, the number of nucleotides effectively targeted) is 10 kb (50 amplicons × 200 nt), which is an order of magnitude shorter than the commercially available HER2 probes currently used in diagnostics. Hybridization with this HER2 probe was specific on both human lymphocyte metaphase spreads as well as in human mammary epithelial (HME) cells processed to preserve their three-dimensional nuclear structure

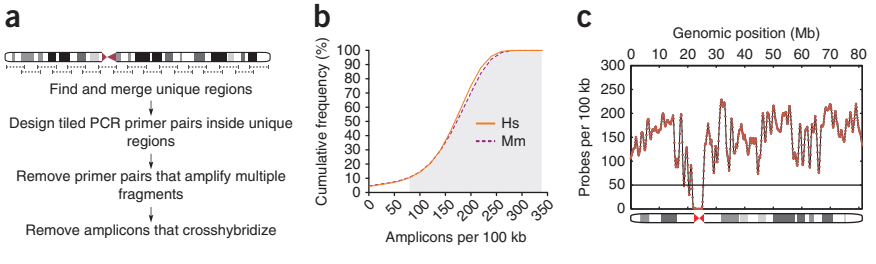

Figure 1 | HD-FISH probe design and synthesis. (**a**) Construction of a database of unique PCR primer pairs and amplicons in the human and mouse genomes. (**b**) Cumulative frequency of amplicons along the human (Hs) and mouse (Mm) genome. In both cases, 90% of the 100-kb tiled windows in which the genome is arbitrarily binned contain ≥80 amplicons (gray highlight). (**c**) Example of amplicon density along human chromosome 17.

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [4]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]Hubrecht Institute–KNAW (Royal Netherlands Academy of Arts and Sciences), Utrecht, The Netherlands. [6]University Medical Center Utrecht, Utrecht, The Netherlands. [7]These authors contributed equally to this work. Correspondence should be addressed to A.v.O. (a.vanoudenaarden@hubrecht.eu).

RECEIVED 29 AUGUST; ACCEPTED 14 NOVEMBER; PUBLISHED ONLINE 23 DECEMBER 2012; DOI:10.1038/NMETH.2306

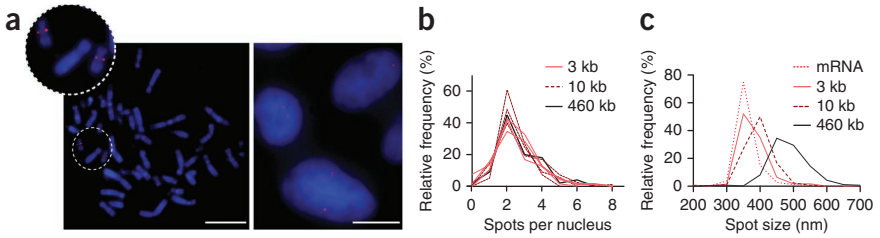

Figure 2 | Specificity and sensitivity of HD-FISH. (**a**) Human *HER2* locus (red) visualized in metaphase spreads (left) and HME cells (right) using a 10-kb ETS probe. Blue, DAPI. Scale bars, 10 µm; inset, 2.3× magnification. (**b**) Distributions of spot counts for three loci on chromosome 17, including *HER2*, visualized with 10-kb compared to 3-kb ETS probes and a HER2 commercial probe spanning 460 kb. (**c**) Distribution of spot sizes for the same HER2 probe as in **a**. mRNA, diffraction-limited HER2 mRNA molecules detected by smRNA FISH.

(using three-dimensional FISH)[8]. Eighty-five percent of interphase nuclei contained between two and four loci, as can be expected in dividing diploid cells (**Fig. 2a**). We obtained similar count distributions for *HER2* and other loci on chromosome 17 over a broad range of ETSs down to 3 kb, demonstrating the sensitivity of our method even with probes derived from only 15 amplicons (**Fig. 2b** and **Supplementary Fig. 3**). Notably, a probe against HER2 with a 3-kb ETS yielded signals with a size (mean diameter, 362 ± 31 nm (mean ± s.d.)) that was similar to the diffraction limit of the microscope used (**Fig. 2c**). For this reason, we named our method HD-FISH.

To demonstrate the scalability of our method, we simultaneously targeted multiple loci spaced evenly on chromosomes 1 and 17 using probes labeled with two alternating fluorophores (**Supplementary Fig. 4a**). Much as with single-locus probes, this 'spotting' strategy was specific on both metaphase spreads as well as on HME interphase cell nuclei after three-dimensional HD-FISH, allowing for the precise enumeration of targeted chromosomes (**Fig. 3a**,**b**, **Supplementary Fig. 4b** and **Supplementary Video 1**). The numbers of spots detected with the first and second fluorophore in the same cell were highly concordant (**Supplementary Fig. 4c**,**d**), indicating that the type of fluorophore does not influence the HD-FISH hybridization efficiency. Notably, our chromosome-spotting strategy was substantially more rapid and cost effective compared to alternative available methods (**Supplementary Fig. 2a**,**b** and **Supplementary Note**).

In the interphase nuclei, HD-FISH spotting yielded clusters of variable size, shape and density that were reminiscent of the chromosomal territories visualized with paint probes[9]. To investigate the spatial relationship between chromosome territories and HD-FISH spot clusters, we performed simultaneous hybridization in HME cells using HD-FISH spotting and paint probes against chromosomes 1 and 17. By visual inspection, HD-FISH spot clusters overlapped largely with the chromosome territories, further emphasizing the specificity of our approach (**Fig. 3c**, **Supplementary Fig. 4e** and **Supplementary Videos 1** and **2**). Taking advantage of the high optical resolution of the HD-FISH signals, we next performed three-dimensional triangulation of the clustered HD-FISH spots for chromosome 17, demonstrating that with probes spaced evenly every 5 Mb, the nucleus occupancy volume of the HD-FISH clusters and the corresponding chromosome territories visualized with paint probes resulted in similar volumes (**Fig. 3d**). In many cases, however, we detected isolated HD-FISH spots in regions with low-intensity paint signals on chromosomes 1 and 17, possibly reflecting better sampling of chromosome volumes (**Fig. 3c** and **Supplementary Fig. 4b**,**e**,**f**).

To extend further the versatility of our method, we combined it with single-molecule RNA FISH (smRNA FISH)[10] in HME cells, simultaneously detecting individual *HER2* loci, sites of active transcription and mature HER2 transcripts (**Fig. 3e**). The number of HER2 transcripts scaled proportionally to the number of *HER2* loci (with a mean of 10.4 HER2 mRNA counts for two DNA loci compared to 20.1 HER2 mRNA counts for four DNA loci; Kolmogorov-Smirnov test $P = 3 \times 10^{-45}$), which is in agreement with previous observations[11] (**Fig. 3f** and **Supplementary Fig. 5**). We also used HD-FISH and smRNA FISH in combination on formalin-fixed paraffin-embedded breast cancer tissue sections. In HER2-positive breast cancers, we found *HER2*

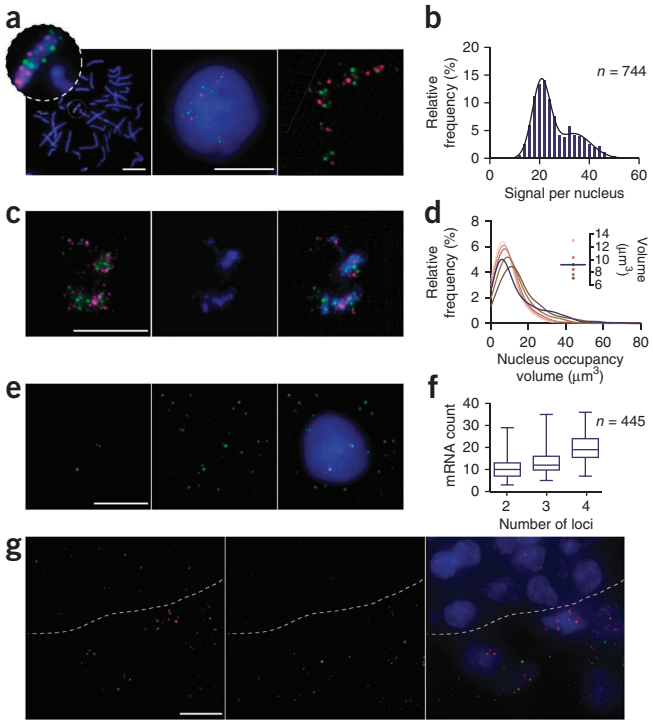

Figure 3 | Versatility of HD-FISH. (**a**) Chromosome 17 spotting with ten alternatively labeled HD-FISH probes in metaphase spreads (left) and HME cells (middle, with a three-dimensional rendering on the right). (**b**) Spot quantification of the HME cells in **a**. (**c**) Chromosome 17 spotting with 16 HD-FISH probes (green and magenta) and paint probes (blue) (left and middle, Z projections; right, three-dimensional rendering). (**d**) Chromosome 17 volume estimation using spotting (purple) compared to paint signals at different thresholds (brown gradient). Inset, range of median values for the curves on the left. Purple line, spotting signal median volume. (**e**,**f**) Visualization (**e**) and quantification (**f**) of *HER2* loci (magenta) and transcripts (green) in HME cells. (**g**) *HER2* loci (magenta) and transcripts (green) in breast cancer stroma (above dashed white lines) compared to tumor cells (below dashed white lines). The *n* values indicate the number of cells analyzed. Scale bars, 10 µm; inset (**a**), 5.5× magnification.

amplification coupled to high HER2 mRNA expression in tumor areas but not in intratumoral stroma, confirming the specificity of our probes (**Fig. 3g**).

HD-FISH can reliably operate at the resolution limit of conventional optical microscopy because of the systematic design of short, unique probes that enable visualization of virtually any human or mouse genomic locus. Diffraction-limited HD-FISH spots are treated as digital instead of analog signals, allowing a threshold-insensitive quantification that is more robust than that obtained using existing methods. Our method is logistically simple, rapid and cost effective and is therefore especially relevant for research laboratories that do not use FISH techniques routinely but wish to study loci for which no probe is commercially available. Our method enables a flexibility that is unachievable with current technologies: because unique primers are synthesized in 96-well plates and PCR reactions are performed separately, different probes consisting of subsets of unique amplicons can be easily and rapidly obtained and combinatorially labeled from the same set of primers. Flexible combinatorial labeling of HD-FISH spotting probes will enable the study of chromosome organization *in situ* by determining precisely the relative position of reference points within spotting clusters. Combining high-definition DNA and RNA FISH methods allows systematic *in situ* analyses of the association between chromosome organization and gene expression, which has not been possible so far because of a lack of robust single-cell quantitative assays. HD-FISH is a powerful method for research and diagnostics with broad applications ranging from the diagnosis of chromosomal aberrations to chromosome architecture studies.

Note added in proof: While the galley proof of this article was in preparation, Beliveau et al.[12] *reported a method for genome-scale generation of DNA FISH probes using microarray oligonucleotide libraries.*

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank P. Junker and S. Semrau for helpful discussions. We are grateful to R.A. Weinberg (Massachusetts Institute of Technology) for providing hTERT-HME1 cells. This work was supported by the US National Institutes of Health (NIH)/National Cancer Institute Physical Sciences Oncology Center at Massachusetts Institute of Technology (U54CA143874), an NIH Pioneer award (1DP1OD003936) and a Nederlandse Organisatie voor Wetenschappelijk Onderzoek Vici award to A.v.O. M.B. and S.I. are sponsored by the Human Frontiers Science Program.

AUTHOR CONTRIBUTIONS

N.C. and A.v.O. conceived the methods. M.B. and N.C. performed experiments, analyzed the data and wrote the manuscript. L.T. generated the genome-wide primer databases, designed the probes and wrote the manuscript. S.K. and S.I. developed software for image processing, provided suggestions on data analysis and corrected the manuscript. A.v.O. guided experiments and data analysis, and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.2306. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Springer Protocols. *Fluorescence In Situ Hybridization (FISH)–Application Guide* (ed. Liehr, T.) (Springer, 2008).
2. Springer Protocols. *Fluorescence In Situ Hybridization (FISH): Protocols and Applications (Methods in Molecular Biology)*. (eds Bridger, J.M. & Volpi, E.V.) (Humana Press, 2010).
3. Boyle, S., Rodesch, M.J., Halvensleben, H.A., Jeddeloh, J.A. & Bickmore, W.A. *Chromosome Res.* **19**, 901–909 (2011).
4. Yamada, N.A. *et al. Cytogenet. Genome Res.* **132**, 248–254 (2011).
5. Rogan, P.K., Cazcarro, P.M. & Knoll, J.H. *Genome Res.* **11**, 1086–1094 (2001).
6. Navin, N. *et al. Bioinformatics* **22**, 2437–2438 (2006).
7. Wiegant, J.C. *et al. Cytogenet. Cell Genet.* **87**, 47–52 (1999).
8. Solovei, I. & Cremer, M. *Methods Mol. Biol.* **659**, 117–126 (2010).
9. Cremer, T. & Cremer, M. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
10. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. & Tyagi, S. *Nat. Methods* **5**, 877–879 (2008).
11. Tsafrir, D. *et al. Cancer Res.* **66**, 2129–2137 (2006).
12. Beliveau, B.J. *et al. Proc. Natl. Acad. Sci. USA* published online, doi:10.1073/pnas.1213818110 (2012).

ONLINE METHODS

Genome-wide primer design. Each chromosome was split *in silico* into 500-nt tiled fragments with 100-nt sliding steps. Fragments were aligned to the genome of interest (hg19/GRCh37 assembly for *Homo sapiens* and NCBIm37/mm9 assembly for *Mus musculus*) using BLAT (http://genome.ucsc.edu/FAQ/FAQblat.html) set to find matches in the genome with 80% or higher similarity[13]. Overlapping unique fragments were merged together. Next, Primer3 (http://primer3.sourceforge.net/) in default mode was run on merged unique DNA sequences to design tiled primer pairs delimiting 200- to 220-nt amplicons[14]. To avoid cross amplification, each primer pair was screened for specificity using electronic PCR (http://www.ncbi.nlm.nih.gov/projects/e-pcr/) with 0- to 2-nt mismatches allowed per primer[15]. Primer pairs yielding non-self products were discarded. Amplicons delimited by unique primer pairs were individually screened for specificity by another round of BLAT searching against the genome of interest. Amplicons with 70% or higher similarity outside their locus were discarded.

Generation of HD-FISH probes. For each probe, forward and reverse primers were synthesized in corresponding wells of separate 96-well plates (stock plates). Forward and reverse primer pairs were mixed and diluted to 5 µM in nuclease-free water using PCR plates (clear LightCycler 480 Multiwell Plate 96, Roche) (dilution plates) at well positions matching the positions of the primers in the stock plate. For each probe, real-time PCR reactions were carried out in the same type of plates (experimental plates) by transferring the appropriate volume of the 5 µM primer dilution from the dilution plate to the corresponding well in the experimental plate. For each reaction (in one well), the following reagents were mixed: 25 µl KAPA SYBR FAST qPCR Master Mix 2×, 4 µl of the 5 µM forward and reverse primers dilution, 1 µl human genomic DNA (diluted 50 ng/µl in Tris-EDTA (TE) buffer, pH 8) and 20 µl nuclease-free water. For each plate, 30 cycles of real-time PCR reactions were performed in a LightCycler 480 instrument (Roche) using the default SYBRGreen program with T_a = 55 °C. After PCR, the contents of all the wells corresponding to a given probe were pooled together in a sterile cell culture basin (VWR) and aliquoted for subsequent ethanol precipitation. Wells in which either no product was observed or the amplification kinetics were substantially different than those in all other wells were excluded.

For labeling, a volume corresponding to 1 µg of each probe was lyophilized and then labeled with the fluorophore of interest using the corresponding ULYSIS Nucleic Acid Labeling Kit (Invitrogen) according to the manufacturer's instructions. Unbound dyes were removed by gel filtration using KREApure columns following the manufacturer's instructions (Kreatech). Labeled probes were stored at −20 °C and were stable at this temperature for up to 6 months.

Reagents. hTERT-HME1 mammary epithelium cells were kindly provided by R.A. Weinberg (Massachusetts Institute of Technology). Metaphase spreads derived from human lymphocytes were obtained from Abbott Diagnostics (30-806010). For the HER2 combined RNA and DNA FISH in breast cancer, HER2 immunohistochemistry control tissue arrays with ten cores were purchased from US Biomax.

The following hybridization buffers were used: (i) metaphase spreads: 1.7× SSC, 50% formamide, 50 mM Na_2HPO_4/NaH_2PO_4, 10% dextran sulfate and 5× Denhardt's solution, pH 7.5; (ii) interphase cells and tissue: 1.7× SSC, 70% formamide, 50 mM Na_2HPO_4/NaH_2PO_4, 10% dextran sulfate and 5× Denhardt's solution, pH 7.5; and (iii) RNA FISH: 25% formamide, 2× SSC, 10% dextran sulfate, 1 mg/ml *Escherichia coli* tRNA, 0.2% BSA and 20 mM ribonucleoside vanadyl complex (RVC). The RNA wash buffer contained 2× SSC and 25% formamide. For microscopy, samples were covered with a mounting solution containing 2× SSC buffer, 10 mM Tris, 0.4% glucose, 100 µg/ml catalase, 37 µg/ml glucose oxidase and 2 mM Trolox.

The commercial HER2 probe (KI-10701) shown in **Figure 2b**,**c** and the chromosomes 1 (KI-30001) and 17 (KI-30017) paint probes were obtained from Kreatech Diagnostics. The HER2 RNA FISH probe comprised 48 different 20-nt oligonucleotide sequences, each with their 3′ end covalently bound to an amino group and purchased from Biosearch Technologies. The probe was coupled to AlexaFluor594 (Invitrogen) as previously described[10].

HD-FISH. For each hybridization, 20 ng of HD-FISH probe (unless specified otherwise) were ethanol precipitated using 20 µg of glycogen as carrier and dissolved in 10, 20 or 30 µl (for metaphase spreads, cells or tissue, respectively) of the appropriate hybridization buffer. For chromosome spotting, 20 ng per hybridization (for chromosome 17, spotting probes comprising ten probes) or 5 ng per hybridization (for chromosomes 17 and 1, spotting probes comprising 16 and 22 probes, respectively) of each probe were precipitated and dissolved in the appropriate hybridization buffer. For chromosome painting alone, paint probes were resuspended in whole chromosome buffer (KI-WCB, Kreatech Diagnostics) according to the manufacturer's instructions. For chromosome painting combined with HD-FISH spotting, 4 µl of paint probe per hybridization were mixed with 16 µl of spotting hybridization mix per hybridization (prepared as described above).

Metaphase spread slides were processed following Abbott Diagnostics' instructions. The desired probe (20 ng for single-locus probes and 6 ng of each probe for chromosome 17 spotting with ten probes) was denatured at 80 °C for 5 min, after which it was applied to a slide. After a 48-h incubation at 37 °C, they were washed following the manufacturer's instructions.

The three-dimensional HD-FISH procedure on interphase nuclei was adapted from ref. 8. When combined with RNA FISH, all steps were identical to those previously described except that starting from fixation, exceptional care was taken to avoid RNase contamination. After DNA hybridization, cover glasses were transferred onto 100 µl of HER2 RNA FISH hybridization solution and incubated at 30 °C for 3h. The cover glasses were then washed twice at 30 °C for 30 min in the RNA wash buffer (the second wash included 20 ng/ml DAPI). After washes, the cover glasses were rinsed with 2× SSC and covered with mounting solution for imaging.

For HD-FISH combined with RNA FISH on breast cancer tissue, tissue arrays were deparaffinized in D-limonene (VWR), rehydrated, post-fixed for 10 min in 4% formalin in 1× PBS, heated for 45 min at 80 °C in 0.01 M sodium citrate, pH 6, and then treated with 0.025% pepsin in 0.01 M HCl. Autofluorescence

was reduced by immersing tissue sections in 1% $NaBH_4$ in 1× PBS solution for 10 min at room temperature. After three washes of 10 min each in RNase-free water, cover glasses were transferred into 2× SSC with 50% formamide, pH 7, and incubated overnight at room temperature. The next day, they were placed into hybridization solution containing 200 ng of HER2 probe, sealed and placed at 37 °C for 3 h before being denatured at 85 °C for 5 min. Hybridization was carried on for 40 h at 37 °C. All steps afterwards were identical to those for three-dimensional HD-FISH combined with RNA FISH.

Microscopy and image processing. All images were acquired at 100× magnification (oil immersion, high numerical aperture Nikon objective) on an inverted epi-fluorescence microscope (Nikon) equipped with a high-resolution charge-coupled device (CCD) camera (Pixis, Princeton Instruments) and controlled by MetaMorph software. DNA spots were identified by thresholding local background–subtracted images. The local background for each nuclear pixel was calculated as either the average signal in the nuclear region in which the pixel appeared or the average signal within the annulus delimited by the first and second minima of an Airy pattern generated by a point source centered at the given pixel. Thresholding was performed on either raw pixel intensities or the integrated signal intensity, defined as the area under a two-dimensional Gaussian surface approximation to the measured Airy pattern. These methods achieved similar results.

The sizes of the DNA spots were defined as the full width at half maximum of a two-dimensional Gaussian approximation to the measured Airy pattern. The diffraction limit was taken to be the full width at half maximum of the Airy pattern generated by smRNA FISH, as previously described[16]. All Gaussian parameters were obtained by maximum likelihood estimation. Three-dimensional animations of HD-FISH spotting and paint probe signals were produced using Imaris v7.2.0 software. Analysis of RNA FISH data was done as previously described[10,16]. Nuclei were manually segmented and dilated by 10 pixels (1.25 µm) to obtain the estimated cell boundaries. The custom code for image processing and analysis was developed in Matlab and Python and will be freely provided on request.

Statistical analyses. The error in HD-FISH spots counting with the chromosome 17 spotting probe (**Fig. 3c**) was estimated as the difference between the expected (ten) and observed number of spots in the AlexaFluor 594 and 647 channels separately. The Bowley skewness of the distribution of single-cell differences between error estimates in the two channels (**Supplementary Fig. 2**) was computed as described at http://mathworld.wolfram.com/BowleySkewness.html using Matlab.

To estimate the volume of the chromosome territories visualized by the paint probes, images were deconvolved using Huygens Essential software. These deconvolved images were then background subtracted and thresholded using custom Python scripts. For each manually segmented cell, the size of the labeled chromosomal volumes was calculated as the number of nuclear pixels (scaled by the physical size of each pixel) with intensity above a given threshold. A range of threshold values was selected by visual inspection, and the single-cell volume analysis was repeated for each value within this range.

We estimated the nucleus occupancy by HD-FISH spot clusters by computing the convex hull (http://mathworld.wolfram.com/ConvexHull.html) of spots clustered using the *k*-means clustering algorithm (http://mathworld.wolfram.com/K-MeansClusteringAlgorithm.html) implemented in Matlab. For each cell, 1,000 iterations of *k*-means were performed with $k = 2$ and random sampling of seed coordinates within the range of dot coordinates. At every iteration, the volumes of the convex hulls of each cluster were added up, and the mean of the summed convex hull volumes over all the iterations was calculated. The mean of clustered dot silhouette values was computed at every iteration, and for each cell, the mean silhouette value over all the iterations was calculated.

13. Kent, W.J. *Genome Res.* **12**, 656–664 (2002).
14. Untergasser, A. *et al. Nucleic Acids Res.* **40**, e115 (2012).
15. Wheeler, D.L. *et al. Nucleic Acids Res.* **31**, 28–33 (2003).
16. Itzkovitz, S. *et al. Nat. Cell Biol.* **14**, 106–114 (2012).

doi:10.1038/nmeth.2306